PENDEKATAN *SMALL AREA ESTIMATION* PADA *SCAN STATISTIC* UNTUK PENDETEKSIAN KANTONG KEMISKINAN

(Adaptation Of Small Area Estimation To Scan StatisticFor Detection Hotspot Poverty)

Reny Ari Noviyanti [1], Ismaini Zain [2]
[1] Mahasiswa Pasca Sarjana, Jurusan Statistika FMIPA, ITS Surabaya
[2] Dosen Pembimbing, Jurusan Statistika FMIPA, ITS Surabaya
Jln. Arief Rahman Hakim, Surabaya, Jawa Timur, 60111
noviyanti.renyari@gmail.com

Abstrak

Dalam rangka mengimplementasikan berbagai program pengentasan kemiskinan diperlukan adanya informasi daerah yang merupakan kantong kemiskinan.Salah satu metode yang bisa digunakan untuk mengidentifikasi kantong kemiskinan adalah Scan Statistic. Permasalahannya, untuk mendeteksi kantong kemiskinan pada level wilayah kecil (kecamatan) diperlukan informasi lengkap dari data populasi sedangkan data kemiskinan pada wilayah kecil tersebut tidak tersedia. Oleh karena itu digunakan metode *Small Area Estimation* (SAE) untuk mendapatkan data kemiskinan pada wilayah kecil (kecamatan) sebagai input dalam *Scan Statistic*. Metode SAE yang digunakan untuk estimasi proporsi kemiskinan level kecamatan adalah *Empirical Bayes* (EB). Untuk mendeteksi kantong kemiskinan digunakan metode *Circular Spatial Scan Statistic*.Hasil *Scan Statististic* berbasis SAE EB diperoleh 6 (enam) kelompok wilayah yang merupakan kantong kemiskinan di Kepulauan Nias.

Kata Kunci:*Empirical Bayes,kantongkemiskinan, SAE, Scan Statistic.*

Abstract

In order to implement poverty alleviation programs, its needed information area which becomes hotspot poverty. One of the methods that can be used to identify hotspot poverty is by using Scan Statistic. The problem, to detect hotspot poverty at the small areas level (sub-districts) are required complete information of population data, while poverty data in small areas is not available. Therefore used Small Area Estimation (SAE) methods to obtain poverty data in small areas (sub-district) as an input for Scan Statistic. SAE method used to estimated poverty proportion in sub-district using Empirical Bayes (EB) method. Circular Spatial Scan Statistic method used to detect hotspot poverty. Result of Scan Statististic using SAE EB obtained 6 (six) groups sub-district which becomes hotspot poverty in Nias.

Keywords: Empirical Bayes, hotspot poverty, SAE, Scan Statistic.

1.PENDAHULUAN

Pengentasan kemiskinan merupakan tantangan global terbesar yang dihadapi dunia dewasa ini dan menjadi syarat mutlak bagi pembangunan berkelanjutan.Agar kebijakan tepat sasaran maka untuk mengimplementasikan berbagai program pengentasan kemiskinan diperlukan adanya informasi daerah yang merupakan kantong kemiskinan.Salah satu metode yang bisa digunakan untuk mengidentifikasi kantong kemiskinan adalah *Scan Statistic*.

Scan Statistic merupakan metode dalam geoinformatika yang mempertimbangkan unsur spasial untuk mendeteksi dan mengevaluasi pengelompokan daerah (*cluster*) yang mempunyai potensi tinggi atau rendah dari suatu kejadian, serta menguji parameter proporsi satu daerah dibandingkan dengan proporsi daerah lain secara [1].

Permasalahannya, untuk mendeteksi kantong kemiskinan pada level wilayah kecil (kecamatan) dengan menggunakan *Scan Statistic* diperlukan informasi lengkap dari data populasi sedangkan data kemiskinan pada wilayah kecil tersebut tidak tersedia. Selama ini data kemiskinan yang dihitung oleh Badan Pusat Statistik (BPS) hanya tersedia sampai level kabupaten/kota yang disebabkan oleh keterbatasan jumlah sampel Survei Sosial Ekonomi Nasional (Susenas) sebagai dasar penghitungan kemiskinan. Estimasi langsung pada wilayah yang lebih kecil tidak bisa dilakukan karena ukuran sampel pada area umumnya sangat kecil sehingga statistik yang diperoleh akan memiliki varian yang besar. Selain itu, ada beberapa area yang tidak terpilih sebagai sampel. Sehingga kesimpulan yang dihasilkan tidak dapat menggambarkan keadaan yang sebenarnya

Apabila dilakukan penambahan jumlah sampel maka akan terkendala oleh biaya dan tenaga yang cukup besar serta waktu yang relatif lama. Upaya lain yang bisa dilakukan adalah mengoptimalkan data yang tersedia dengan menggunakan metode *Small Area Estimation* (SAE) untuk mendapatkan data kemiskinan pada wilayah kecil (kecamatan) sebagai input dalam *Scan Statistic*.

Metode SAE yang akan digunakan untuk memperoleh estimator sebagai input *Scan Statistic* dalam penelitian ini adalah *Empirical Bayes* (EB). Kelebihan metode *Empirical Bayes* yang dikatakan oleh Rao [2] bahwa penduga SAE EB mempunyai *mean square error* yang kecil.

Berdasarkan yang telah diuraikan terlihat bahwa dengan mengaplikasikan SAE pada *Scan Statistic* akan memberikan penghematan biaya, waktu dan tenaga dalam menentukan daerah kantong kemiskinan karena penelitian tidak perlu dilakukan untuk seluruh populasi, tetapi cukup dengan memanfaatkan data atau informasi administrasi yang tersedia.

Untuk mendeteksi kantong kemiskinan digunakan metode *Circular Spatial Scan Statistic* karena berdasarkan perbandingan dari beberapa metode, *Circular Spatial Scan Statistic* mempunyai kemampuan lebih tinggi dibandingkan dengan metode lainnya saat *cluster* yang dideteksi adalah *circular* [3]. Oleh karena itu, dalam penelitian ini digunakan metode *Circular Spatial Scan Statistic* berbasis *Small Area EstimationEmpirical Bayes* untuk mendeteksi kantong kemiskinan (*hotspot poverty*).

2.TINJAUAN PUSTAKA

2.1 Small Area Estimation (SAE)

Small Area Estimation (SAE) merupakan metode estimasi tidak langsung (*indirect estimation*) yang mengkombinasikan antara data survei dengan data pendukung lain misalnya dari data sensus sebelumnya yang memuat variabel dengan karakteristik yang sama dengan data survei sehingga dapat digunakan untuk menduga area yang lebih kecil dan memberikan tingkat akurasi yang lebih baik [2]. Model *small area* dikelompokkan menjadi dua jenis yaitu:

2.2.1 *Small AreaEstimation* Berbasis Area

Pada model *Small Area Estimation* berbasis area, data pendukung yang tersedia hanya sampai level area. Model level area menghubungkan penduga langsung *small area* dengan data pendukung dari domain lain untuk setiap area.

Parameter *small area* yang ingin diamati adalah θ_i. Model linier yang menjelaskan hubungan tersebut adalah:

$$\theta_i = x_i^T \beta + z_i v_i \qquad i = 1, 2 \ldots.., m \qquad (1)$$

dengan

$\beta = (\beta_1, \ldots, \beta_p)^T$ adalah koefisien regresi berukuran $p\, x 1$

$z_i =$ konstanta positif yang diketahui

$v_i =$ pengaruh acak *small area,* diasumsikan $v_i \sim iid\, N(0, \sigma_v^2)$

Dalam membuat kesimpulan tentang populasi diasumsikan bahwa nilai estimasi langsung $\hat{\theta}_i$ diketahui maka dapat dinyatakan sebagai berikut:

$$\hat{\theta}_i = \theta_i + e_i \qquad i = 1, 2 \ldots.., m \qquad (2)$$

dimana e_i adalah *sampling error,* diasumsikan $e_i \sim iid\, N(0, \psi_i)$.

Model SAE untuk level area terdiri dari dua tingkat komponen model yaitu komponen model estimasi tidak langsung sesuai dengan persamaan (1) dan komponen model estimasi langsung sesuai persamaan (2). Model pada persamaan (1) dan (2) jika digabung akan membentuk persamaan sebagai berikut:

$$\hat{\theta}_i = x_i^T \beta + z_i v_i + e_i \qquad i = 1, 2 \ldots.., m \qquad (3)$$

2.2.2 *Small AreaEstimation* Berbasis Unit

Pada model *Small Area Estimation* berbasis unit diasumsikan bahwa data variabel penyerta unit $x_{ij}^T = (x_{ij1}, x_{ij2}, \ldots\ldots, x_{ijp})^T$ tersedia untuk setiap elemen ke- j pada area ke-i namun kadang cukup dengan rata-rata populasi $\bar{x}_{ij}$ diketahui saja. Selanjutnya variabel respon y_{ij} diasumsikan berkaitan dengan x_{ij} sehingga bentuk persamaan model SAE berbasis unit adalah:

$$y_{ij} = x_{ij}^T \beta + e_i + u_{ij}, \quad j = 1, 2, \ldots, n_i \quad i = 1, 2, \ldots, m \qquad (4)$$

Dimana e_i diasumsikan sebagai variabel random yang *iid*, sedangkan $u_{ij} = c_{ij} \tilde{u}_{ij}$ dengan c_{ij} adalah konstanta dan $\tilde{u}_{ij}$ adalah variabel random yang *iid* dan bebas terhadap e_i dimana $E(\tilde{u}_{ij}) = 0$ dan $Var(\tilde{u}_{ij}) = \sigma_u^2$. Seringkali diasumsikan bahwa e_i dan u_{ij} berdistribusi normal.

2.2 Small Area Estimation Empirical Bayes (SAE EB)

Pada metode pendugaan langsung (*direct estimation*) dalam kerangka sampel yang diambil dalam setiap area lokal ke-*i*, pendugaan proporsi (p_i) dapat ditulis sebagai berikut:

$$p_i = \frac{\sum_j y_{ij}}{n_i} \tag{5}$$

dimana

y_{ij} = nilai nol atau satu, tergantung dari apakah individu ke-*j* pada area lokal ke-*i* memenuhi karakteristik tertentu yang ingin diperhatikan.

n_i = ukuran sampel dari area lokal ke-*i*.

i = *1, 2, ..., m*, dengan *m* adalah banyaknya kecamatan yang terpilih sebagai *small area.*

Pada pendugaan proporsi melalui pendekatan *Empirical Bayes*, jika suatu variabel respon merupakan variabel kategorik biner dimana y_{ij} = 1 atau 0 serta semua data kovariat x_{ij} yang berasosiasi dengan y_{ij} tersedia untuk semua area, yaitu $x_{ij} = x_i$, maka proporsi dari sampel dapat ditransformasi menggunakan transformasi *arcsin*, dan mereduksi model menjadi model level area. Transformasi *arcsin* digunakan untuk menstabilkan nilai varian dari estimasi langsung [2].

$$\hat{y}_i = \arcsin\sqrt{\hat{p}_i} \tag{6}$$

diasumsikan $\hat{y}_i \sim iid\left(y_i, \psi_i\right)$ dimana $V\left(\hat{y}_i\right) \approx G_i \big/ \left(4n_i\right)$, $\hat{G}_i = \hat{V}_i\left[\hat{p}_i\left(1-\hat{p}_i\right) \big/ n_i\right]^{-1}$ dan $V_i = V\left(\hat{p}_i\right)$ merupakan varian dari $\hat{p}_i$.

Data dari variabel pendukung (*auxiliary variables*) diikutsertakan dalam model. Data pendukung yang tersedia hanya sampai pada level area yaitu $x_i^T = \left(x_{i1}, \ldots, x_{ip}\right)^T$ maka model untuk pendekatan *Empirical Bayes* dengan menggunakan model pada persamaan (3) dimana $v_i \sim iid\ N\left(0, \sigma_v^2\right)$ dan $e_i \sim iid\ N\left(0, \psi_i\right)$, v_i dan e_i saling bebas, β dan σ_v^2 tidak diketahui sedangkan ψ_i diasumsikan diketahui. Misal σ_v^2 dan ψ_i disimbolkan dengan A dan D_i, selanjutnya merupakan estimator Bayes untuk θ_i, dengan mengikuti model Bayes sebagai berikut:

(i) $y_i|\theta_i \sim N(\theta_i, D_i)$

(ii) $\theta_i \sim N(x_i^T\beta, A)$ adalah sebaran prior untuk θ_i, $i = 1, 2, \cdots, m$

Berdasarkan kaidah model bayes maka diperoleh estimator bayes untuk θ_i adalah:

$$\hat{\theta}_i^B = E(\theta_i|y_i, \beta, A) = x_i^T\beta + (1 - B_i)(y_i - x_i^T\beta) \qquad (7)$$

dimana $B_i = D_i \big/ (A + D_i)$

$$MSE(\hat{\theta}_i^B) = Var(\theta_i|y_i, \beta, A) = AD_i + (A + D_i) \qquad (8)$$

Ketika parameter A diketahui, maka β pada formula diatas dapat diestimasi dengan metode *Maximum Likelihood*. Oleh karena A tidak diketahui, sehingga untuk estimasi parameter A adalah dengan menggunakan metode *Maximum Likelihood Estimation (MLE)* atau *Restricted/Residual Maximum Likelihood (REML)*. Meskipun estimator A terdapat pelanggaran asumsi kenormalan, akan tetapi pendugaan dengan menggunakan metode REML tetap menghasilkan estimator A yang konsisten [4]. Sehingga A dan β bisa diestimasi dan diperoleh penduga *Empirical Bayes* sebagai berikut:

$$\hat{\theta}_i^{EB} = x_i^T\hat{\beta} + (1 - \hat{B}_i)(y_i - x_i^T\hat{\beta}) \qquad (9)$$

dimana $\hat{B}_i = D_i \big/ (\hat{A} + D_i)$

Estimator MSE *Empirical Bayes* adalah sebagai berikut:

$$MSE(\hat{\theta}_i^{EB}) = Var(\theta_i|y_i, \hat{\beta}, \hat{A}) = \hat{A}D_i + (\hat{A} + D_i) \qquad (10)$$

Oleh karena adanya estimasi pada nilai A dan β sehingga menyebabkan estimator MSE menjadi *underestimate*. Hal tersebut dapat dikoreksi dengan menggunakan pendekatan *jackknife*.Langkah-langkah pendekatan *Jackknife* dalam menduga MSE dugaan *Empirical Bayes* jika $MSE(\hat{\theta}_i^B) = AD_i + (A + i) = g_{1i}(A)$ dimana A diduga oleh s_v^2 adalah sebagai berikut [5].

1. Hitung nilai h_{1i} dengan rumus:

$$h_{1i} = g_{1i}\left(s_v^2\right) - \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[g_{1i}\left(s_{v(-u)}^2\right) - g_{1i}\left(s_v^2\right)\right] \quad (11)$$

$g_{1i}\left(s_{v(-u)}^2\right)$ diperoleh dengan menghapus pengamatan ke-*u* pada himpunan data $g_{1i}\left(s_v^2\right)$.

2. Hitung nilai h_{2i} dengan rumus:

$$h_{2i} = \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[\left(\hat{\theta}_{i(-u)}^{EB}\right) - \left(\hat{\theta}_i^{EB}\right)\right]^2 \quad (12)$$

dimana $\left(\hat{\theta}_{i(-u)}^{EB}\right)$ diperoleh dengan menghapus pengamatan ke-*u* pada himpunan data $\left(\hat{\theta}_i^{EB}\right)$

3. Hitung nilai MSE:

$$MSE\left(\hat{\theta}_i^{EB}\right) = h_{1i} + h_{2i} \quad (13)$$

2.3 Metode *Scan Statistic*

Metode *Scan Statistic* adalah suatu metode statistika untuk mendeteksi gerombol (*hotspot*) dalam suatu wilayah yang signifikan secara statistik terhadap resiko kasus tertentu.Gerombol dari *hotspot* ditentukan dengan aturan bahwa area dalam gerombol tersebut memiliki resiko relatif lebih tinggi dibanding yang lainnya [6].Penentuan *hotspot* atau *most likely cluster* (MLC) pada *Scan Statistic* didasarkan pada uji rasio kemungkinan (*log likelihood ratio*/LLR).Untuk menentukan *cluster*, pertama kali dibangun *zone* (*scanning window*) dimana setiap *zone* merupakan kandidat untuk menjadi MLC [1].

Algoritma yang digunakan untuk menentukan MLC yaitu [7]:

1. Menentukan daerah yang akan diteliti.
2. Menentukan data spasial untuk setiap lokasi.
3. Membentuk kumpulan *scanning window*. Setiap *scanning window* merupakan kandidat dari MLC.
4. Membentuk hipotesis H_0dan H_1.
5. Membangun *Log Likelihood Ratio*(LLR)berdasarkan H_0 dan H_1.
6. Menghitung *Log Likelihood Ratio* untuk setiap *scanning window*.
7. Mencari daerah yang potensial, dimana daerah potensial merupakan *scanning window* dengan nilai LLR tertinggi
8. Melakukan pengujian hipotesis dengan cara menggunakan pengujian hipotesis Monte Carlo.

3.METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini bersumber dari Susenas 2011, Podes 2011, dan data spasial berupa peta wilayah administrasi dan koordinat kantorkecamatan hasil Pemetaan Sensus Penduduk 2010.

Variabel respon dalam penelitian ini adalah proporsikemiskinan $\left(\hat{\theta}_i\right)$ level kecamatan.Variabel penyerta $\left(x_i\right)$ yang digunakan adalah sebagai berikut.

x_1 : Persentase keluarga yang tinggal di bantaran sungai.

x_2 : Rasio lembaga pendidikan (SD, SMP, SMA, SMEA) terhadap total penduduk.

x_3 : Rasio sarana kesehatan terhadap total penduduk.

x_4 : Rasio tenaga kesehatan terhadap total penduduk.

x_5 : Persentase keluarga pertanian.

x_6 : Persentase penduduk penerima Jamkesmas.

x_7 : Persentase SKTM yang dikeluarkan.

Software yang digunakan untuk mengolah data dalam penelitian adalah SPSS, Excel, SAS, Matlab,SatScan dan ArcView.

Tahapan analisis adalah sebagai berikut:

1. Menduga proporsi kemiskinanuntuk masing-masing kecamatan di Kepulauan Nias dengan menggunakan metode pendugaan langsung (*direct estimation*).
2. Eksplorasi data untuk melihat keterkaitan antara variabel respon proporsi kemiskinandengan keseluruhan variabel penyerta (*auxiliary variables*).
3. Membentuk model SAE metode *Empirical Bayes*berbasis area untuk mengestimasi proporsi kemiskinansetiapkecamatan ke-*i* dengan menggunakan informasi dari variabel penyerta.

4. Menghitung *Mean Square Error* (*MSE*) dari hasil pendugaan proporsi kemiskinankecamatan ke-*i* metode *Empirical Bayes* dengan menggunakan pendekatan metode *Jackknife*.
5. Membandingkan MSE pendugaan langsung (*direct estimate*) dengan pendugaan SAE *Empirical Bayes* pendekatan metode *Jacknife*.
6. Menggunakan data proporsi kemiskinanlevelkecamatan di Kepulauan Niasyang diperoleh dari hasil pendugaan SAE *Empirical Bayes*sebagai populasi. Selain itu, mengambil data spasial yang berupa titik koordinat masing-masing kantorkecamatan.
7. Mengidentifikasi kandidat *cluster/hotspot*dengan algoritma yang ada pada*Scan Statistic*.
8. Menghitung jumlah populasi dan jumlah kasus (penduduk miskin) untuk setiap *cluster*.
9. Menghitung *log likelihood ratio* dari setiap *cluster*kemudian menguji signifikansi *cluster*dengan simulasi Monte Carlo untuk mendapatkan *p-value*.
10. Menghitung *Relative Risk*(RR) untuk setiap *cluster*.
11. Mengintepretasikan hasil dengan membuat deskripsi kantong kemiskinan sehingga diperoleh prioritas lokasi pengentasan kemiskinanlevel kecamatan di KepulauanNiasberdasarkan *cluster*yang signifikan pada $\alpha = 0,05$.

4.HASIL DAN PEMBAHASAN

Pada penelitian ini, model *Scan Statistic* yang digunakan adalah model bernoulli. Fungsi probabilitas yang menyatakan probabilitas banyaknya kejadian m_{y_i} dalam *sub-region* ke-*i*adalah :

$$f\left(m_{y_i}\right) = \begin{cases} \begin{pmatrix} N_{y_i} \\ m_{y_i} \end{pmatrix} p^{m_{y_i}} \left(1-p\right)^{N_{y_i}-m_{y_i}} , y_i \in Z \\ \begin{pmatrix} N_{y_i} \\ m_{y_i} \end{pmatrix} q^{m_{y_i}} \left(1-q\right)^{N_{y_i}-m_{y_i}} , y_i \notin Z \end{cases} \tag{14}$$

dengan*i=1,2,...,k*; Z merupakan region atau daerah; N_{y_i} merupakan banyak kejadian dalam populasi. Fungsi likelihood yang terbentuk adalah:

$$L(p,q;Z) = \prod_{y_i \in Z} \begin{pmatrix} N_{y_i} \\ m_{y_i} \end{pmatrix} p^{m_{y_i}} (1-p)^{N_{y_i}-m_{y_i}} \prod_{y_i \notin Z} \begin{pmatrix} N_{y_i} \\ m_{y_i} \end{pmatrix} q^{m_{y_i}} (1-q)^{N_{y_i}-m_{y_i}} \tag{15}$$

Fungsi likelihood *scanning window* untuk setiap pasangan (m_Z,N_Z) adalah:

$$L(Z,p,q) = p^{m_Z}(1-p)^{N_Z-m_Z}\ q^{m_G-m_Z}(1-q)^{(N_G-N_Z)-(m_G-m_Z)} \qquad (16)$$

dimana

m_Z = Banyaknya kasus di satu daerah

N_Z = Jumlah populasi di satu daerah

m_G = Total banyak kasus di seluruh daerah

N_G = Total populasi di seluruh daerah

p = Probabilitas individu terkena permasalahan didalam *scanning window*

q= Probabilitas individu terkena permasalahan di luar *scanning window*

Persamaan (16) menyatakan peluang sukses dari satu daerah dikalikan dengan peluang gagal di daerah tersebut dikalikan dengan peluang sukses di luar daerah dikalikan dengan peluang gagal di luar daerah.

Untuk mendeteksi *zone* MLC, ditentukan zone $\hat{z}$ yang memaksimumkan fungsi *likelihood*, dirumuskan sebagai:

$$L(\hat{Z},p,q) = \left(p^{m_Z}(1-p)^{N_Z-m_Z}\right)\left(q^{m_G-m_Z}(1-q)^{(N_G-N_Z)-(m_G-m_Z)}\right)\prod_{y_i\in G}\frac{N_{y_i}!}{(N_{y_i}-m_{y_i})!\,m_{y_i}!} \qquad (17)$$

Ruang parameter dalam hipotesis H_0 adalah suatu himpunan yang anggotanya merupakan parameter dalam fungsi likelihood jika H_0 benar atau tidak ada kasus, yaitu $\omega = \{(p,q): 0 \le p = q \le 1\}$. Ruang parameter dalam hipotesis H_1 yaitu $\Omega = (p,q): 0 \le p \le 1, 0 \le q \le 1,\ Z: p > q\}$.

Fungsi likelihood berdasarkan H_0 benar diperoleh apabila $p = q$ adalah:

$$L(\omega) = p^{m_G}(1-p)^{N_G-m_G}\prod_{y_i\in G}\frac{N_{y_i}!}{(N_{y_i}-m_{y_i})!\,m_{y_i}!} \qquad (18)$$

Fungsi likelihood jika H_1 benar (apabila terdapat MLC) adalah:

$$L(\Omega) = \left(p^{m_Z}(1-p)^{N_Z-m_Z}\right)\left(q^{m_G-m_Z}(1-q)^{(N_G-N_Z)-(m_G-m_Z)}\right)\prod_{y_i\in G}\frac{N_{y_i}!}{(N_{y_i}-m_{y_i})!\,m_{y_i}!} \qquad (19)$$

Dengan menggunakan metode *MLE* didapatkan estimator parameter p dan q sebagai berikut:

Ln fungsi kemungkinan pada kondisi tidak ada kasus yaitu $\ln L(\omega)$ adalah:

$$\ln L(\omega) = m_G \ln p + (N_G - m_G)\ln(1-p) + \prod_{y_i \in G} \ln \frac{N_{y_i}!}{(N_{y_i} - m_{y_i})!\, m_{y_i}!} \quad (20)$$

Untuk memaksimumkan fungsi likelihood didapat apabila turunan pertama terhadap parameter bernilai nol dan turunan kedua negatif, sehingga turunan pertama dari $\ln L(\omega)$ terhadap parameter p adalah:

$$\frac{\partial \ln L(\omega)}{\partial p} = \frac{m_G}{p} - \frac{N_G}{(1-p)} + \frac{m_G}{(1-p)} = 0$$

$$\frac{m_G}{p} = \frac{N_G - m_G}{(1-p)}$$

$$\hat{p} = \frac{m_G}{N_G} \quad (21)$$

Ln fungsi kemungkinan pada kondisi tidak ada kasus yaitu ln $L(\Omega)$ adalah:

$$\ln L(\Omega) = m_Z \ln(p) + N_Z \ln(1-p) - m_Z \ln(1-p) + (m_G - m_Z)$$
$$\ln(q) + N_G \ln(1-q) - N_Z \ln(1-q) + (m_G - m_Z)\ln(1-q)$$
$$+\prod_{y_i \in G} \ln \frac{N_{y_i}!}{(N_{y_i} - m_{y_i})!\, m_{y_i}!} \quad (22)$$

Persamaan akan maksimum apabila turunan parsial pertama terhadap parameter p dan q sama dengan nol. Dicari turunan parsial pertama *ln* $L(\Omega)$ terhadap parameter p:

$$\frac{\partial \ln L(\Omega)}{\partial p} = \frac{m_Z}{p} - \frac{N_Z}{(1-p)} + \frac{m_Z}{(1-p)} = 0$$

$$\frac{m_Z}{p} = \frac{N_Z - m_Z}{(1-p)}$$

$$\hat{p} = \frac{m_Z}{N_Z} \quad (23)$$

Untuk mencari estimasi dari parameter q maka turunan parsial pertama *ln* $L(\Omega)$ terhadap parameter q:

$$\frac{\partial \ln(\Omega)}{\partial q} = \frac{m_G}{q} - \frac{m_Z}{q} - \frac{N_G}{(1-q)} + \frac{N_Z}{(1-q)} - \frac{m_G}{(1-q)} - \frac{m_Z}{(1-q)} = 0 \quad \frac{m_G - m_Z}{q} = \frac{N_G - N_Z + m_G - m_Z}{(1-q)}$$

$$\hat{q} = \frac{m_G - m_Z}{N_G - N_Z} \quad (24)$$

Dengan mensubstitusi $\hat{p}$ dan $\hat{q}$ dalam persamaan fungsi *likelihood* diperoleh:

$$L(\hat{\omega}) = \left(\frac{m_G}{N_G}\right)^{m_G} \left(1 - \frac{m_G}{N_G}\right)^{N_G - m_G} \sum_{y_i \in G} \frac{N_{y_i}!}{(N_{y_i} - m_{y_i})!\, m_{y_i}!} \quad (25)$$

$$L\left(\hat{\Omega}\right)=\left(\left(\frac{m_Z}{N_Z}\right)^{m_Z}\left(1-\frac{m_Z}{N_Z}\right)^{N_Z-m_Z}\right)\left(\left(\frac{m_G-m_Z}{N_G-N_Z}\right)^{m_G-m_Z}\right.$$
$$\left.\left(1-\left(\frac{m_G-m_Z}{N_G-N_Z}\right)\right)^{(N_G-N_Z)-(m_G-m_Z)}\right)\sum_{y_i\in G}\frac{N_{y_i}!}{\left(N_{y_i}-m_{y_i}\right)!} \quad (26)$$

dengan $L\left(\hat{\Omega}\right)=Max_{Z\in\hat{Z}}L\left(\Omega\right)$

Untuk mendeteksi *most likely cluster* yang nilai probabilitas dalam kelompok daerah tersebut lebih tinggi dibanding probabilitas di luar kelompok, *log likelihood ratio* (LLR) dirumuskan sebagai berikut:

$$\nabla=\frac{\left(\left(\frac{m_Z}{N_Z}\right)^{m_Z}\left(1-\frac{m_Z}{N_Z}\right)^{N_Z-m_Z}\right)\left(\left(\frac{m_G-m_Z}{N_G-N_Z}\right)^{m_G-m_Z}\left(1-\left(\frac{m_G-m_Z}{N_G-N_Z}\right)\right)^{(N_G-N_Z)-(m_G-m_Z)}\right)}{\left(\frac{m_G}{N_G}\right)^{m_G}\left(1-\frac{m_G}{N_G}\right)^{N_G-m_G}} \quad (27)$$

Nilai LLR tersebut dihitung untuk setiap *scanning window* yang terbentuk. Berdasarkan kumpulan *scanning window* dan nilai LLR dari masing-masing pasangan (m_Z,N_Z), maka *potential cluster* atau calon *most likely cluster* dapat ditentukan, yaitu *scanning window* yang mempunyai nilai LLR tertinggi. Untuk mengetahui apakah calon *most likely cluster* yang didapat signifikan secara statistik, maka dilakukan pengujian dengan menghitung nilai signifikansi atau *p-value* dengan menggunakan pengujian hipotesis Monte Carlo.

4.1*Small Area Estimation Empirical Bayes* pada *Scan Statistic*

Dengan mengganti pendugaan langsung (*direct estimate*) dengan pendugaan SAE *Empirical Bayes* maka fungsi likelihood termodifikasi menjadi:

$$L\left(\hat{\Omega}\right)=\left(\hat{p}_i^{EB\,m_Z}\left(1-\hat{p}_i^{EB}\right)^{N_Z-m_Z}\right)\left(\hat{q}_i^{EB\,m_G-m_Z}\left(1-\hat{q}_i^{EB}\right)^{(N_G-N_Z)-(m_G-m_Z)}\right)\prod_{y_i\in G}\frac{N_{y_i}!}{\left(N_{y_i}-m_{y_i}\right)!m_{y_i}!}$$

(28)

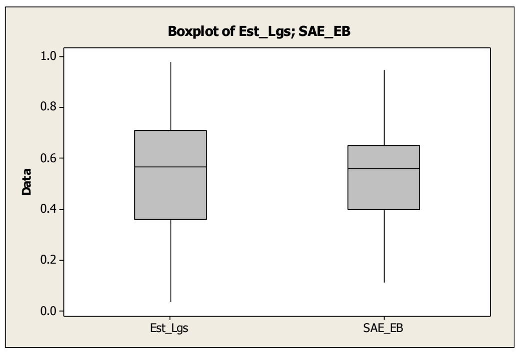
$$L\left(\hat{\omega}\right)=\left(\frac{\sum_{i=1}^{m}\hat{p}_i^{EB}}{m}\right)^{m_G}\left(1-\frac{\sum_{i=1}^{m}\hat{p}_i^{EB}}{m}\right)^{N_G-m_G}\prod_{y_i\in G}\frac{N_{y_i}!}{\left(N_{y_i}-m_{y_i}\right)!m_{y_i}!} \quad (29)$$

Sehingga *log likelihood ratio* (LLR) *Scan Statistic* dengan menggunakan SAE *Empirical Bayes* dirumuskan sebagai berikut:

$$\nabla = \frac{\left(\left(\hat{p}_i^{EB}\right)^{m_Z}\left(1-\hat{p}_i^{EB}\right)^{N_Z-m_Z}\right)\left(\left(\hat{q}_i^{EB}\right)^{m_G-m_Z}\left(1-\left(\hat{q}_i^{EB}\right)\right)^{(N_G-N_Z)-(m_G-m_Z)}\right)}{\left(\frac{\sum_{i=1}^{m}\hat{p}_i^{EB}}{m}\right)^{m_G}\left(1-\frac{\sum_{i=1}^{m}\hat{p}_i^{EB}}{m}\right)^{N_G-m_G}} \quad (30)$$

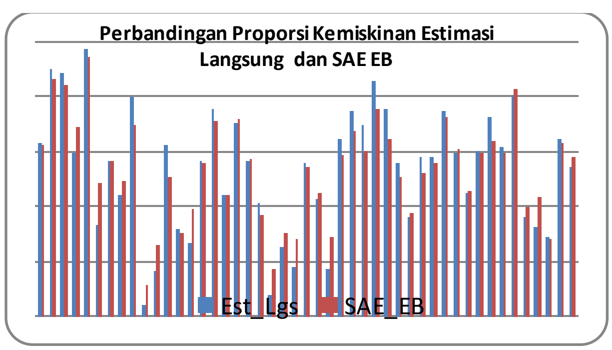
4.2 Penerapan Pada Proporsi Kemiskinan Di Kepulauan Nias

Data proporsi kemiskinan diperoleh dari Susenas 2011 yang diambil dari 47 kecamatan di Kepulauan Nias. Hasil estimasi proporsi kemiskinan berdasarkan *Small Area Estimation Empirical Bayes* diperoleh nilai proporsi kemiskinan mempunyai varian lebih rendah dibandingkan dengan estimasi langsung.



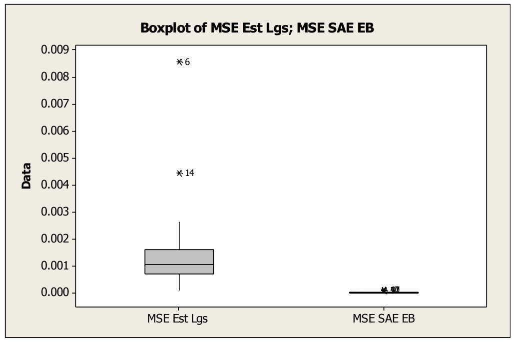
Gambar 1. Boxplot Proporsi Kemiskinan Hasil Estimasi Langsung dan SAE EB

Hasil perbandingan proporsi kemiskinan berdasarkan estimasi langsung dan SAE EB adalah sebagai berikut.



Gambar 2. Perbandingan Proporsi Kemiskinan Hasil Estimasi Langsung dan SAE EB

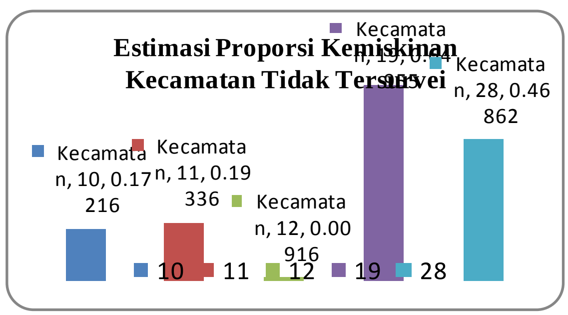
Salah satu ukuran untuk mengetahui seberapa baik estimator parameter yang diperoleh dalam suatu estimasi adalah dengan melihat nilai *Mean Square Error* (MSE) dari estimator yang diperoleh. Perbandingan nilai MSE estimasi langsung dan estimasi hasil SAE EB adalah sebagai berikut.



Gambar 3. Boxplot MSE Proporsi Kemiskinan Hasil Estimasi Langsung dan SAE EB

Berdasarkan gambar 3 menunjukkan bahwa estimasi proporsi kemiskinan dengan menggunakan SAE EB menghasilkan MSE yang lebih kecil dibandingkan estimasi langsung.

Untuk estimasi proporsi kemiskinan di kecamatan yang tidak tersurvei, diperoleh dengan menggunakan metode pendugaan sintetik, dengan asumsi perilaku antar kecamatan di Kepulauan Nias adalah sama (nilai sama). Hasil estimasi adalah sebagai berikut:



Gambar 4. Estimasi Proporsi Kemiskinan

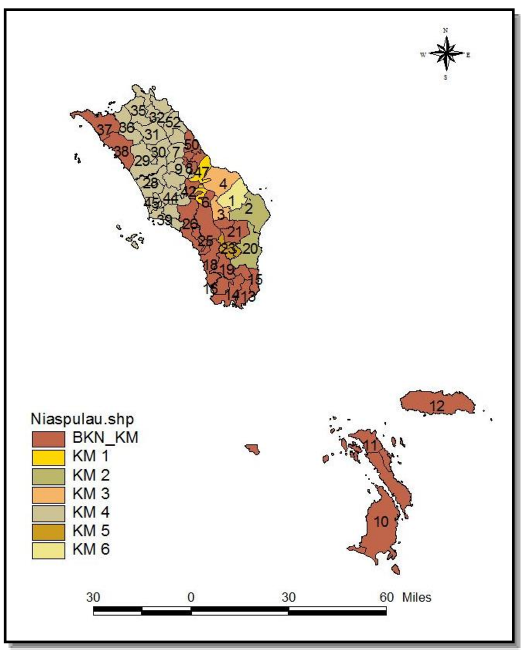
Kecamatan Tidak Tersurvei

Berdasarkan hasil estimasi proporsi kemiskinan dengan menggunakan SAE EB didapatkan proporsi kemiskinan setiap kecamatan di Kepulauan Nias yang selanjutnya digunakan sebagai input dalam *Scan Statistic*. Hasil pendeteksian kantong kemiskinan berdasarkan *Scan Statistic* adalah sebagai berikut:

Tabel 1. Ringkasan Hasil *Scan Statistic*

Kantong Kemiskinan	*Relative Risk* (RR)	LLR	*Monte Carlo Rank*	p-value
1	1.71	17651.13	1/1000	< 0.00
2	1.55	9121.34	1/1000	< 0.00
3	1.41	3722.72	1/1000	< 0.00
4	1.16	1950.81	1/1000	< 0.00
5	1.23	893.05	1/1000	< 0.00
6	1.20	550.22	1/1000	< 0.00

Tabel 1 menunjukkan *scanning window* yang terbentuk dengan melihat nilai *log likelihood ratio* dan uji hipotesis Monte Carlo diperoleh 6 (enam) kelompok wilayah yang merupakan daerah kantong kemiskinan di Kepulauan Nias. Peta kantong kemiskinan di Kepulauan Nias adalah sebagai berikut.



Gambar 5. Peta Kantong Kemiskinan di Kepulauan Nias

Kantong kemiskinan yang memiliki *relative risk* (RR) semakin tinggi menandakan tingkat kemiskinan yang semakin tinggi dan didapatkan dari 6 (enam) *cluster* kantong kemiskinan, daerah *cluster* utamanya yaitu Kecamatan Gunung Sitoli Idanoi dan Ma'u.

5. PENUTUP

5.1 Kesimpulan

Penerapan *Scan Statistic* untuk pendeteksian kantong kemiskinan level kecamatan memerlukan informasi seluruh area dalam populasi. Permasalahannya, data kemiskinan yang bersumber dari Susenas hanya tersedia sampai level kabupaten/kota. *Small Area Estimation Empirical Bayes* (SAE EB) digunakan dalam menyelesaikan persoalan *small area* tersebut, termasuk memprediksi area yang tidak tersurvei. SAE mempunyai sifat statistik (*unbiased*, varians minimum, stabil), dimana sifat tersebut juga dimiliki oleh *Scan Statistic*. Sehingga pendeteksian kantong kemiskinan dengan *Scan Statistic* berbasis SAE EB bisa digunakan untuk menggantikan peran penduga langsung (*direct estimation*).

5.2 Saran

Penentuan variabel penyerta(*auxilary variables*) dan metode estimasi yang digunakan sangat mempengaruhi kualitas dari hasil aplikasi SAE pada *Scan Statistic*. Oleh karena itu perlu eksplorasi yang lebih mendalam untuk menentukan variabel penyerta serta metode SAE yang tepat.

DAFTAR PUSTAKA

[1]Jung I, Kulldorff M, Klassen A, " *A Spatial Scan Statistic for Ordinal Data*," Harvard Medical School and Harvard Pilgrim Health Care; Johns Hopkins Bloomberg School of Public Health, Sponsors : Centers for Disease Control and Prevention (CDC); Association of American Medical Colleges (AAMC), Grant number: MM-0870, 2007.

[2] Rao, J.N.K, "*Small Area Estimation*", John Wiley and Sons, Inc., New York, 2003.

[3] Tango, T. dan Takahashi, K, "A Flexibly Shaped Spatial Scan Statistic For Detecting Clusters", *International Journal of Health Geographics, Volume 4:11*, 2005.

[4] Jiang, J., Lahiri, P., dan Wan. S. M, *"A Unified Jackknife Theory"*, Annals of Statistics, 30, 2007.

[5] Kurnia, A. dan Notodiputro, K.A, "Penggunaan Metode Jackknife dalam Pendugaan Area Kecil", *Makalah disampaikan pada Seminar Nasional Matematika*, UNPAD Bandung, 22 April 2006.

[6]Song, Changhong, and Martin Kulldorff, "Power evaluation of disease clustering tests", *International Journal of Health Geographics*, 2003.

[7] Kulldorff, M, *"A Spatial Scan Statistic"*, Communication in Statistics: Theory and Methods, 26(6), 1481-1496, 1997.

[8] Kuldorff M, Nagarwala N, "*Spatial disease cluster: Detection and inference*", Statistic in Medicine14:799-810, 1995.

[9] Kulldorff, M, "*SaTScan User Guide for version 9.0"*, 2010, http://www.satscan.org/

[10] Siswantining, Titin, "*Geoinformatika Pada Kasus Area Kecil Dan Penerapannya Untuk Mendeteksi Kantong-Kantong Kemiskinan Di Jember",* Disertasi, Bogor : Institut Pertanian Bogor, 2013.